Recent Advances for Maximum-Entropy Sampling

Jon Lee

University of Michigan
Ann Arbor, Michigan

MIP 2022 / DanFest
23–26 May 2022

Differential Entropy and the MESP

$N := \{1, 2, \ldots, n\}$.
Random $\mathcal{Y}_N := (\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_n)^\mathsf{T}$ with continuous density g_N.

Goal: Given $0 < s < n$, choose $S \subset N$, with $|S| = s$, so that observing $\mathcal{Y}_S$ maximizes the "information" obtained about $\mathcal{Y}_N$, where "information" is *differential entropy* $h(S) := -E[\ln g_S(\mathcal{Y}_S)]$, see Shannon 1948.

Some calculations:

- If $\mathcal{Z}_S := A\mathcal{Y}_S + b$, where $A \in \mathbb{R}^{|S|\times|S|}$ is invertible, and $b \in \mathbb{R}^{|S|}$, then $h(\mathcal{Z}_S) = h(\mathcal{Y}_S) + \ln|\det A|$.
  $\Longrightarrow$ differences of entropies are meaningful.
- If $\mathcal{Y}_S$ has a non-degenerate joint Gaussian distribution with covariance matrix $C[S,S]$, then

$$h(\mathcal{Y}_S) = \frac{1}{2}\ln\det(2\pi e\, C[S,S]) = \frac{1}{2}\Big((1+\log(2\pi))|S| + \ln\det C[S,S]\Big).$$

The Constrained Maximum-Entropy Sampling Problem

$$
\begin{aligned}
z(C,s,A,b) := \quad & \max\Big\{ \ln \det C[S,S] \ : \ |S| = s; \\
& \qquad \textstyle\sum_{j\in S} a_{ij} \le b_i,\ i=1,2,\ldots,m \Big\} \\
& = \max\left\{ \ln \det C[S(x),S(x)] \ : \ \mathbf{e}^{\mathsf{T}} x = s, \right. \\
& \qquad \left. Ax \le b,\ x \in \{0,1\}^n \right\},
\end{aligned}
\qquad \text{(CMESP)}
$$

- The term “MESP” comes from the experimental-design literature (Shewry and Wynn 1987). It has been developed in that literature, and more broadly in statistics (Sebastiani and Wynn 2000).
- As a finite discrete problem, statisticians employed simple interchange heuristics, which actually do quite well on instances of modest size. The mathematical-optimization community began to get involved starting with Ko, Lee, and Queyranne 1995.

An application: environmental monitoring I

see Caselton, Kan, and Zidek 1992; Wu and Zidek 1992; Guttorp, Le, Sampson, and Zidek 1993; Brown, Le, and Zidek 1994; Wang, Le, and Zidek 2020; Zidek, Sun, and Le 2000; Lee 2012; Le and Zidek 2006+`EnviroStat`; Al-Thani and Lee 2020+`MESgenCov`.

Setting: We have univariate time-series $\mathcal{Y}_j(t)$ at spatially dispersed locations ("monitoring sites") $j \in N$, for discrete time points $t = 1, 2, \ldots, T$. The *network-contraction* problem is to choose a subset $S \subset N$, with $|S| = s$ (given), with the idea that future observations will only be collected at S. From the data, we can calculate a sample covariance matrix C, and then we formulate an instance of MESP.

Data: The NADP (National Acidic Deposition Program) maintains the NTN (National Trends Network); see NADP 2018). The NTN measures the chemistry of precipitation at 379 monitoring sites, with some (weekly) data available as far back as 1978; at present, 255 sites are active.

An application: environmental monitoring II

Data: Power plants burning fossil fuels produces a large fraction of the (SO_2) emissions in the US. Emissions of nitrogen oxides (NO_x) come from transportation, power plants, and other industrial sources. SO_2 ultimately leads to sulfuric acid (H_2SO_4). Similarly, nitrogen oxides lead to nitric acid (HNO_3). Presence of these acids in significant amounts decreases the pH of precipitation to ≤ 4.0, and this is considered "acid rain", which has adverse impacts on terrestrial and aquatic creatures and on infrastructure.

Toward monitoring for acid rain, the NTN analyzes samples for: hydrogen ion (H^+ measured as pH), sulfate (SO_4^{2-}), and nitrate (NO_3^-). Additional analysis is carried out for ammonium (NH_4^+), chloride (Cl^-), calcium (Ca^{2+}), magnesium (Mg^{2+}), potassium (K^+), ortho-phosphate (PO_4^{3-}). and sodium (Na^+).

An application: environmental monitoring III

Figure: Aerochem Metrics Precipitation Collector

An application: environmental monitoring IV

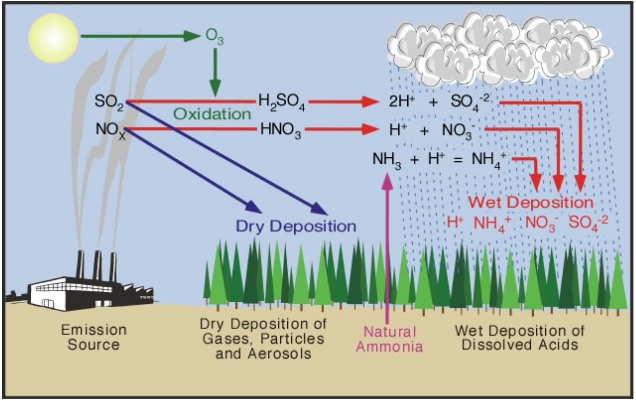

Figure: The "acid rain" process

An application: environmental monitoring V

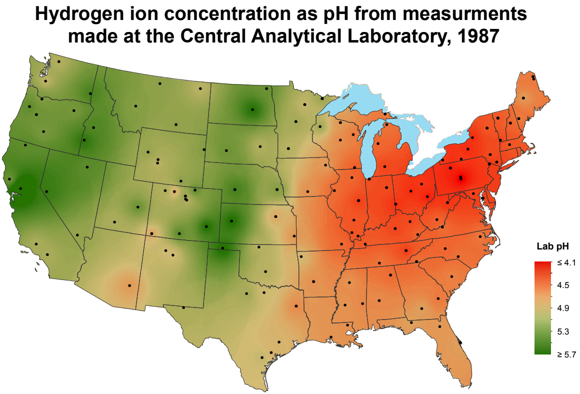

Figure: NTN "acid rain" data: 1987

An application: environmental monitoring VI

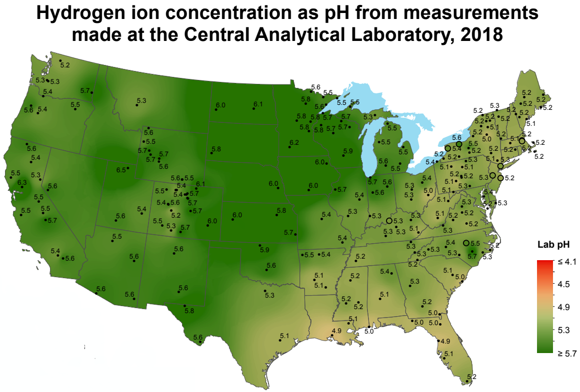

Figure: NTN “acid rain” data: 2018

An application: environmental monitoring VII

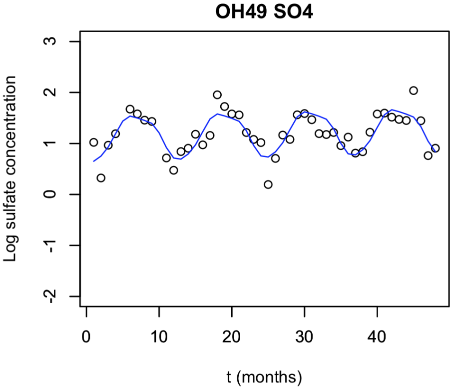

Figure: Log sulfate concentration over a four-year period at a site

Some preliminary facts

- MESP is NP-hard; reduction from: Does a graph G contain a stable set (of vertices) of cardinality s. Just take $C := A(G) + nI_n$ (see Ko, Lee, and Queyranne 1995).
- Special case: If C (or C^{-1}) is tridiagonal, then solvable by dynamic programming, see Al-Thani and Lee 2021.
- $\ln \det C[S, S]$ is a (generally non-monotone) submodular function ($\cong$ the "Hadamard-Fischer inequalities").
- Bounds and bounds:
  - $z(C, s, A, b) = z(\gamma C, s, A, b) - s \ln \gamma$, leading to the equivalent "scaled problem".
  - $z(C, s, A, b) = z(C^{-1}, n - s, -A, b - A\mathbf{e}) + \ln \det C$, leading to the equivalent "complementary problem".
  - $z(C, s, A, b) \leq z(M \circ C, s, A, b)$, for any correlation matrix M (from Oppenheim's inequality), leading to the related "masked problem".

Algorithmic approaches

- A variety of greedy and local-search heuristics for MESP, integer-linear programming heuristics for CMESP, aiming at good lower bounds for large instances (see Ko, Lee, and Queyranne 1995; Lee 1998).
- Approximation algorithms for MESP (and for very restricted CMESP), modulo scaling, mainly based on the big and still-growing literature on submodular maximization, which started with Nemhauser, Wolsey, and Fisher 1978.
- Exact algorithms, mostly based on a "branch-and-bound" framework, aimed at solution of moderate-sized instances.
- Among all continuous random vectors $\mathcal{Y}_S$ with covariance matrix $C[S, S]$, the ones with maximum entropy are Gaussian (from Gibb's inequality).

  $\Longrightarrow$ Upper bounds on Gaussian entropy are upper bounds on entropy.

Branch-and-Bound

- We maintain a list $\mathcal{L}$ of subproblems of the form $\mathsf{L}(F_0, F_1)$ where $F_1 \subset S \subset N \setminus F_0$, and a lower bound LB on $z(C, s, A, b)$.
- Notice that the optimal value of $\mathsf{L}(F_0, F_1)$ is just $\ln \det C[F_1, F_1] + z(C_{F_1}[N \setminus F_0 \setminus F_1, N \setminus F_0 \setminus F_1], s - |F_1|)$.
- Initially, the list $\mathcal{L}$ contains only the given CMESP, and LB could be the value of any heuristic solution.
- For a subproblem $\mathsf{L}(F_0, F_1)$, we consider the continuous relaxation of its feasible region:

$$\big\{x \in \mathbb{R}^n \ : \mathbf{e}^\mathsf{T} x = s,\ Ax \leq b,\ 0 \leq x \leq \mathbf{e},\ x_j = 0 \text{ for } j \in F_0,\ x_j = 1 \text{ for } j \in F_1\big\}.$$

- **Key invariant properties to maintain for branch-and-bound**
  - Every subproblem on $\mathcal{L}$ is feasible, but the continuous relaxation of the corresponding subproblem does not have a unique feasible solution.
  - If there is a feasible solution S of CMESP with $\ln \det C[S, S] >$ LB then S is a feasible solution for some subproblem on $\mathcal{L}$.
- Then we can stop when $\mathcal{L} = \emptyset$, and we will have $LB = z(C, s, A, b)$.

How we process the list $\mathcal{L}$

- An iteration of the algorithm chooses and removes some subproblem $\mathsf{L}(F_0, F_1) \in \mathcal{L}$.
- Next, we apply some upper bounding method to the chosen subproblem $\mathsf{L}(F_0, F_1)$.
- If the calculated upper bound for the subproblem $\mathsf{L}(F_0, F_1)$ is less than or equal to LB, we simply discard $\mathsf{L}(F_0, F_1)$, and the key invariant properties are maintained.
- If this is not the case, we choose a *branching index* $j \in N \setminus F_1 \setminus F_0$. From this, we define:
  - the *in child* as $L(F_1 + j, F_0)$, and
  - the *out child* as $L(F_1, F_0 + j)$.
- It is easy to see that every feasible solution of $\mathsf{L}(F_0, F_1)$ is feasible for either its in child or its out child.

How we treat our children

- If $|F_1 + j| = s$, then $F_1 + j$ is the unique set S satisfying $F_1 + j \subset S \subset N \setminus F_0$, $|S| = s$; so in this case, we discard the in-child, and if $\sum_{k \in F_1+j} A[\cdot, k] \leq b$, then we update LB $:= \max\{\text{LB}, \ln \det C[F_1 + j, F_1 + j]\}$.
- If $|F_0 + j| = n - s$, then $N \setminus (F_0 + j)$ is the unique set S satisfying $F_1 \subset S \subset N \setminus (F_0 + j)$, $|S| = s$; so in this case, we discard the out-child, and if $\sum_{k \in N \setminus (F_0+j)} A[\cdot, k] \leq b$, then we update LB $:= \max\{\text{LB}, \ln \det C[N \setminus (F_0 + j), N \setminus (F_0 + j)]\}$.
- If a child cannot be discarded based on one of the these rules above, we go further. We consider the feasible region of the continuous relaxation of the CMESP associated with a child. With a single linear-optimization problem, we can determine a maximal set of linear equations satisfied by all points of the continuous relaxation (see Freund, Roundy, and Todd 1985). If this is a single point $\hat{x}$, then we discard the child, and if $\hat{x}$ is binary, then we update LB $:= \max\{\text{LB}, \ln \det C[S(\hat{x}), S(\hat{x})]\}$.

Spectral upper bounds

For MESP, we have the *spectral bound*:

Proposition (see Ko, Lee, and Queyranne 1995)

$$z(C,s) \le \sum_{\ell=1}^{s} \ln \lambda_\ell(C),$$

and with the mask $M = I_n$, we have

Proposition (see Hoffman, Lee, and Williams 2001)

$$z(C,s) \le \sum_{\ell=1}^{s} \ln \operatorname{diag}(C)_{[\ell]},$$

Spectral upper bounds, continued

- Extension to CMESP via Lagrangian relaxation: Lee 1998.
- Much more on masking the spectral bound
  - locally optimizing the mask: Anstreicher and Lee 2004.
  - same but with a Cholesky-factor variable for the mask: Burer and Lee 2007.
  - combinatorial masks: Hoffman, Lee, and Williams 2001.
  - combinatorial masks, matching, and integer-linear optimization: Lee and Williams 2003.
  - tridiagonal masks: Al-Thani and Lee 2021.

Convex-programming relaxations: general advantages

- Easy passage from MESP to CMESP (i.e., no conceptual difficulty in including our constraints $Ax \leq b$).
- Variable fixing methodology for "convex MINLP" (based on duality) can be directly and effectively applied in cases where the constraints $\mathbf{0} \leq x \leq \mathbf{e}$ are explicit:

Theorem

Let

- *LB be the objective-function value of a feasible solution for CMESP,*
- $(\cdot, \hat{\upsilon}, \hat{\nu})$ *be a feasible solution for the dual of our convex relaxation with objective-function value* $\hat{\zeta}$*.*

Then, for every optimal solution x^* *for CMESP, we have:*

$$x_k^* = 0, \ \forall\ k \in N \ \textit{such that } \hat{\zeta} - \mathit{LB} < \hat{\upsilon}_k \ ,$$
$$x_k^* = 1, \ \forall\ k \in N \ \textit{such that } \hat{\zeta} - \mathit{LB} < \hat{\nu}_k \ .$$

A warm-up for the "NLP bound"

Let $x(S)$ be the characteristic vector of S. We have

$$\operatorname{Diag}(x(S))C\operatorname{Diag}(x(S)) + \operatorname{Diag}(\mathbf{e} - x(S)) = \left(\begin{array}{c|c} C[S,S] & 0 \\ \hline 0 & I \end{array}\right).$$

Leading to the following formulation of CMESP:

$$\begin{array}{rl} \max & \left\{ \ln\det\left(\operatorname{Diag}(x)C\operatorname{Diag}(x) + \operatorname{Diag}(\mathbf{e}-x)\right) \; : \right. \\ & \left. \mathbf{e}^{\mathsf{T}}x = s,\; Ax \le b,\; x \in \{0,1\}^n \right\}, \end{array}$$

but the objective function is not concave on $x \in [0,1]^n$.

NLP bound: Anstreicher, Fampa, Lee, and Williams 1999

The first convex-programming upper bound for CMESP.

$$z_{\mathsf{NLP}}(C,s,A,b) := \max\left\{ \ln\,\det R(x) \;:\; \mathbf{e}^{\mathsf{T}}x = s,\; Ax \le b,\; 0 \le x \le \mathbf{e} \right\},$$

where

$$R(x) := \mathrm{Diag}(x^{p/2})C\,\mathrm{Diag}(x^{p/2}) + \mathrm{Diag}(d_i^{x_i} - d_i x_i^{p_i} \;:\; i \in N).$$

Properties (see Chen, Fampa, and Lee 2021b):

- With appropriate choices of parameters, the NLP bound is a convex program, smooth, and efficiently solvable with custom (and probably general-purpose) nonlinear-programming software (e.g., Knitro).
- Complementing and scaling can be effectively employed; see Anstreicher, Fampa, Lee, and Williams 1999

linx bound: Anstreicher 2020

A bound of the form $\log\det($linear in x$)$.

$$z_{\text{linx}}(C,s,A,b) := \max\left\{\tfrac{1}{2}\ln\det L(x)\ :\ \mathbf{e}^{\mathsf{T}}x = s;\ Ax \le b;\ 0 \le x \le \mathbf{e}\right\},$$

where $L(x) := C\,\mathrm{Diag}(x)C + \mathrm{Diag}(\mathbf{e}-x)$.

Nice properties:

- Convex, smooth, and efficiently solvable with nonlinear-programming software (e.g., Knitro and SDPT3); see Chen, Fampa, and Lee 2021b.
- Self-complementary, like the spectral bound; see Anstreicher 2020.
- Bound is very sensitive to scaling, but finding the optimal scale factor can be cast as a univariate convex minimization problem; see Chen, Fampa, Lambert, and Lee 2021a.

Justifying the linx bound

Let $x(S)$ be the characteristic vector of S, and let $T := N \setminus S$. We have

$$C\,\mathrm{Diag}(x(S))C + \mathrm{Diag}(\mathbf{e} - x(S)) = \left(\begin{array}{c|c} C[S,S]^2 & C[S,S][S,T] \\ \hline C[S,T]^{\mathsf{T}}C[S,S] & C[S,T]^{\mathsf{T}}C[S,T] + I \end{array}\right).$$

If $C[S,S]$ is singular, then we can see that $L(x(S))$ is singular. So we have that $f(x(S)) = \ln\,\det C[S,S] = -\infty$ in this case.

If $C[S,S]$ is nonsingular. Employing the Schur complement determinant formula and taking logarithms, we have

$$\begin{aligned} &\ln\,\det(C\,\mathrm{Diag}(x(S))C + \mathrm{Diag}(\mathbf{e} - x(S))) \\ &= 2\ln\,\det C[S,S] + \\ &\quad \ln\,\det(C[S,T]^{\mathsf{T}}C[S,T] + I - C[S,T]^{\mathsf{T}}C[S,S]C[S,S]^{-2}C[S,S]C[S,T]) \\ &= 2\ln\,\det C[S,S]. \end{aligned}$$

Masking the linx bound: Chen, Fampa, and Lee 2022

We have the scaled and masked linx bound

$$\mathrm{linx}(C,s;M,\gamma) := \max\{f(C,s;M,\gamma;x) \,:\, \mathbf{e}^{\mathsf{T}}x = s,\ 0 \le x \le \mathbf{e}\},$$

where $f(C,s;M,\gamma;x) :=$

$$\tfrac{1}{2}\left(\ln\,\det\left(\gamma(C\circ M)\,\mathrm{Diag}(x)(C\circ M) + \mathrm{Diag}(\mathbf{e}-x)\right) - s\log\gamma\right).$$

Theorem (Masking can help linx a lot, even under optimal scaling)

There is an infinite sequence of positive-semidefinite matrices $\{C_n\}_{n\in 4\mathbb{Z}_{++}}$, such that

$$\min_{\gamma>0} \mathrm{linx}\left(C_n, \tfrac{n}{2}; J, \gamma\right) - \min_{\bar{\gamma}>0} \mathrm{linx}\left(C_n, \tfrac{n}{2}; I, \bar{\gamma}\right) \ge bn$$

for some positive scalar $b \ge 0.024036$.

Factorization bound: Nikolov 2015

Also see: Li and Xie 2020; Chen, Fampa, and Lee 2021b.

$$z_{\mathsf{Fact}}(C,s,A,b;F) := \max\left\{\textstyle\sum_{\ell=1}^{s}\log\left(\lambda_\ell(F^{\mathsf{T}}\,\mathrm{Diag}(x)F)\right) \;:\; \mathbf{e}^{\mathsf{T}}x=s,\ Ax\le b,\ 0\le x\le \mathbf{e}\right\},$$

where $C=FF^{\mathsf{T}}$.

This is *not* a convex program, but its Lagrangian dual is:

$$\begin{array}{ll} z_{\mathsf{DFact}}(C,s,A,b;F) := & \\ \quad \min\ -\sum_{\ell=k-s+1}^{k}\log\left(\lambda_\ell\left(\Theta\right)\right)+\nu^{\mathsf{T}}\mathbf{e}+\pi^{\mathsf{T}}b+\tau s-s & \\ \qquad \text{subject to:} & \text{(DFact)} \\ \qquad \mathrm{diag}(F\Theta F^{\mathsf{T}})+\upsilon-\nu-A^{\mathsf{T}}\pi-\tau\mathbf{e}=0, & \\ \qquad \Theta\succ 0,\ \upsilon\ge 0,\ \nu\ge 0,\ \pi\ge 0. & \end{array}$$

Factorization bound, continued

And taking a further Lagrangian dual, we can get to the even more tractable:

$$z_{\mathsf{DDFact}}(C, s, A, b; F) := \max\left\{\Gamma_s(F(x)) \;:\; \mathbf{e}^\mathsf{T} x = s,\ Ax \le b,\ 0 \le x \le \mathbf{e}\right\}. \qquad \text{(DDFact)}$$

Nice properties (see Chen, Fampa, and Lee 2021b):

- Convex, reasonably smooth, and efficiently solvable with nonlinear-programming software (e.g., Knitro).
- Factorization bound is independent of the factorization $C = FF^\mathsf{T}$.
- Factorization bound is invariant under scaling.
- Factorization bound provably dominates the spectral bound.
- Complementation technique is useful for the factorization bound.

But what is Γ_s?

Factorization bound, continued

Lemma (Nikolov 2015)

Let $\lambda \in \mathbb{R}^k_+$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ and let $0 < s \leq k$. There exists a unique integer ι, with $0 \leq \iota < s$, such that

$$\lambda_\iota > \tfrac{1}{s-\iota} \textstyle\sum_{\ell=\iota+1}^{k} \lambda_\ell \geq \lambda_{\iota+1},$$

with the convention $\lambda_0 = +\infty$.

With the hypotheses of the lemma, let ι be the unique integer above. We define

$$\phi_s(\lambda) := \textstyle\sum_{\ell=1}^{\iota} \log\left(\lambda_\ell\right) + (s-\iota)\log\left(\tfrac{1}{s-\iota}\textstyle\sum_{\ell=\iota+1}^{k} \lambda_\ell\right).$$

Next, for $X \in \mathbb{S}^k_+$, we define $\Gamma_s(X) := \phi_s(\lambda(X))$.

BQP bound: Anstreicher 2018; Helmberg 1995 (unpub.)

The idea now is to lift $x \in \mathbb{R}^n$ to a matrix variable $X \in \mathbb{S}^n$, relaxing the nonconvex constraint $X = xx^\mathsf{T}$.

$$\begin{aligned} z_{\mathsf{BQP}}(C, s, A, b) := \max \{ \ln \det M(x, X) \ : \ & (x, X) \in \mathbb{R}^n \times \mathbb{S}^n, \\ & Ax \leq b,\ \mathbf{e}^\mathsf{T} x = s, \\ & X - xx^\mathsf{T} \succeq 0,\ \operatorname{diag}(X) = x,\ X\mathbf{e} = sx \}, \end{aligned}$$

where $M(x, X) := C \circ X + \operatorname{Diag}(\mathbf{e} - x)$.

Nice properties:

- Convex, smooth, and efficiently solvable with nonlinear-programming software (e.g., SDPT3); see Anstreicher 2018.
- Bound is sensitive to scaling, but finding the optimal scale factor can be cast as a univariate convex-minimization problem (see Chen, Fampa, Lambert, and Lee 2021a).
- We can do variable fixing, even though the constraints $0 \leq x \leq \mathbf{e}$ are implicit from $X - xx^\mathsf{T} \succeq 0$ and $x = \operatorname{diag}(X)$; see Anstreicher 2018.

Mixing bounds: Chen, Fampa, Lambert, and Lee 2021a

- We consider $m \geq 1$ convex relaxations for CMESP, indexed by $i = 1, \ldots, m$:

$$v_i := \max\left\{ f_i(L_i(x)) \ : \ \mathbf{e}^\mathsf{T} x = s,\ Ax \leq b,\ 0 \leq x \leq \mathbf{e} \right\},$$

  where, for $i = 1, \ldots, m$, $k_i \leq n$, $L_i : \mathbb{R}^n \to \mathbb{S}_+^{k_i}$ are affine functions, and $f_i : \mathbb{S}_+^{k_i} \to \mathbb{R}$ are concave functions. We write $L_i(x) := L_{i0} + L_{i1}x_1 + \cdots + L_{in}x_n$ and $L_{ij} \in \mathbb{S}^{k_i}$, for $i = 1, \ldots, m$ and $j = 0, \ldots, n$. We note that the objective functions of DDFact, comp-DDFact, and linx can be written as $f_i(L_i(x))$.
- For a "weight vector" $\alpha \in \mathbb{R}_+^m$, such that $\mathbf{e}^\mathsf{T}\alpha = 1$, we define the *mixing bound* (see Chen et al. 2021a for a more general setting):

$$v(\alpha) := \max\left\{ \textstyle\sum_{i=1}^m \alpha_i f_i(L_i(x)) \ : \ \mathbf{e}^\mathsf{T} x = s,\ Ax \leq b,\ 0 \leq x \leq \mathbf{e} \right\}.$$

- The goal is to minimize the mixing bound over α (and any parameters for the individual bounds).

Mixing bounds, continued

- We have convexity in the weight vector α.
- We have an algorithmic framework to minimize over α, and we can indeed get improvements on unmixed bounds, whenever unmixed bounds are close to one another.
- We have a methodology to fix variables using duality; this is where we exploit the form $f_i(L_i(x))$.
- We have ways to improve on the mixing of the BQP bound and the comp-BQP bound, via cuts.

Finally

Where are the computational results?

Finally

Where are the computational results?

In the papers! (see references at the end of this slide deck)

Finally

Where are the computational results?

In the papers! (see references at the end of this slide deck)

Where can I read more?

Finally

Where are the computational results?

In the papers! (see references at the end of this slide deck)

Where can I read more? In our book:

Maximum-Entropy Sampling:

Algorithms and application

Marcia Fampa and Jon Lee

Forthcoming in 2022 (Springer, ORFE Series)

Thanks!

Questions?

References I

Al-Thani H, Lee J (2020) An R package for generating covariance matrices for maximum-entropy sampling from precipitation chemistry data. *SN Operations Research Forum* Volume 1:Article 17 (21 pages).

Al-Thani H, Lee J (2021) Tridiagonal MESP and tridiagonal masks. *LAGOS 2021 proceedings, Procedia Computer Science* 195:127–134.

Anstreicher KM (2018) Maximum-entropy sampling and the Boolean quadric polytope. *Journal of Global Optimization* 72(4):603–618.

Anstreicher KM (2020) Efficient solution of maximum-entropy sampling problems. *Operations Research* 68(6):1826–1835.

Anstreicher KM, Fampa M, Lee J, Williams J (1999) Using continuous nonlinear relaxations to solve constrained maximum-entropy sampling problems. *Mathematical Programming, Series A* 85(2):221–240.

Anstreicher KM, Lee J (2004) A masked spectral bound for maximum-entropy sampling. *mODa 7—Advances in Model-Oriented Design and Analysis*, 1–12, Contrib. Statist. (Physica, Heidelberg).

References II

Brown PJ, Le ND, Zidek JV (1994) Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics. La Revue Canadienne de Statistique* 22(4):489–509.

Burer S, Lee J (2007) Solving maximum-entropy sampling problems using factored masks. *Mathematical Programming, Series B* 109(2–3):263–281.

Caselton WF, Kan L, Zidek JV (1992) Quality data network designs based on entropy. Walden A, Guttorp P, eds., *Statistics in the environmental and earth sciences*, 10–38, London: Elsevier Science (London: Elsevier Science).

Chen Z, Fampa M, Lambert A, Lee J (2021a) Mixing convex-optimization bounds for maximum-entropy sampling. *Mathematical Programming, Series B* 188:539–568.

Chen Z, Fampa M, Lee J (2021b) On computing with some convex relaxations for the maximum-entropy sampling problem. Preprint at: `http://arxiv.org/abs/2112.14291`.

References III

Chen Z, Fampa M, Lee J (2022) Masking Anstreicher's linx bound for improved entropy bounds. *Operations Research (to appear)* .

Freund RM, Roundy R, Todd MJ (1985) Identifying the set of always-active constraints in a system of linear inequalities by a single linear program. Sloan W.P. No. 1674-85 (Rev). Available at: `https://dspace.mit.edu/handle/1721.1/2111`.

Guttorp P, Le ND, Sampson PD, Zidek JV (1993) Using entropy in the redesign of an environmental monitoring network. Patil G, Rao C, Ross N, eds., *Multivariate Environmental Statistics*, volume 6, 175–202 (North-Holland).

Hoffman A, Lee J, Williams J (2001) New upper bounds for maximum-entropy sampling. *mODa 6—Advances in Model-Oriented Design and Analysis (Puchberg/Schneeberg, 2001)*, 143–153, Contrib. Statist. (Physica, Heidelberg).

Ko CW, Lee J, Queyranne M (1995) An exact algorithm for maximum entropy sampling. *Operations Research* 43(4):684–691.

References IV

Le ND, Zidek JV (2006) *Statistical Analysis of Environmental Space-Time Processes*. Springer Series in Statistics (Springer, New York).

Lee J (1998) Constrained maximum-entropy sampling. *Operations Research* 46(5):655–664.

Lee J (2012) Maximum entropy sampling. El-Shaarawi A, Piegorsch W, eds., *Encyclopedia of Environmetrics, 2nd ed.*, 1570–1574 (Wiley).

Lee J, Williams J (2003) A linear integer programming bound for maximum-entropy sampling. *Mathematical Programming, Series B* 94(2–3):247–256.

Li Y, Xie W (2020) Best principal submatrix selection for the maximum entropy sampling problem: Scalable algorithms and performance guarantees. Preprint at: `https://arxiv.org/abs/2001.08537`.

NADP (2018) National Acidic Deposition Program, National Trends Network, `https://nadp.slh.wisc.edu/ntn/`.

References V

Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions. I. *Mathematical Programming* 14(3):265–294.

Nikolov A (2015) Randomized rounding for the largest simplex problem. *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, 861–870, STOC '15 (New York, NY, USA: Association for Computing Machinery).

Sebastiani P, Wynn HP (2000) Maximum entropy sampling and optimal Bayesian experimental design. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62(1):145–157.

Shannon CE (1948) A mathematical theory of communication. *The Bell System Technical Journal* 27(3):379–423.

Shewry MC, Wynn HP (1987) Maximum entropy sampling. *Journal of Applied Statistics* 46:165–170.

References VI

Wang Y, Le ND, Zidek JV (2020) Approximately optimal spatial design: How good is it? *Spatial Statistics* 37:100409, frontiers in Spatial and Spatio-temporal Research.

Wu S, Zidek JV (1992) An entropy-based analysis of data from selected nadp/ntn network sites for 1983–1986. *Atmospheric Environment. Part A. General Topics* 26(11):2089 – 2103.

Zidek JV, Sun W, Le ND (2000) Designing and integrating composite networks for monitoring multivariate Gaussian pollution fields. *Journal of the Royal Statistical Society. Series C. Applied Statistics* 49(1):63–79.